State Estimation in Low Voltage Networks Using Smart Meters: Statistical Analysis of the Errors

Pablo Arboleya, Dept. of Electrical Engineering
University of Oviedo
Gijón, Spain

***Abstract*—In the present work, the assumptions made by the classical state estimation theory based on weighted least squares are revised when this theory is applied to low voltage systems with non-aggregated smart meters data. The measurement errors will be analysed obtaining some interesting conclusions about their probability density functions. Finally, these conclusions will be validated by means of different power flow tests using different smart meters set-up.**

***Index Terms*—State estimation, smart-meters, low voltage systems, quasi-static time series, distribution systems, power flow, smart grid, smart meters, DSSE.**

I. Introduction

The massive installation of Smart Meters (SMs) in low voltage distribution networks is a fact. By 2020, the European Commission estimates that more than 200 million devices will be installed [1]. In some countries like Spain, by December 2018, all domestic consumptions should be monitored using SMs [2].

Nowadays, most of the distribution companies are using this technology just for billing purposes. However, researchers are proposing new applications for taking advantage of the large amount of data and information provided by these devices. These applications go from client segmentation or other kind of data extraction like theft detection [3] to topology observation [4], [5], faults detection [6] or phase identification [7], [8].

One of the most studied and referenced application is the state estimation of the distribution grid using the SMs data [9]–[14]. The approaches to the problem are different, but it could be concluded that most of the researchers agree that conventional estate estimation techniques applied to transmission systems must be updated and modified in order to be applied to the distribution level [9].

Many efforts has been employed in the development of new techniques and methodologies, but in some cases, some assumptions of the classical state estimation theory have been adopted as canons. It is very common to find cases using the classical theory and assuming that the measurement errors are normally distributed. For instance in [10], a method for phase connectivity verification and estimation is proposed. The authors can detect which phases are connected to the distribution grid using an algorithm based on a three phase state estimation procedure with equality and conditional constraints. The minimisation function is the classical function used by the weighted least squares (WLS) method and the proposed algorithm combines real-time measurements with data provided by the smart meters sampled at a low rate (15min to 1hour).

In [12], the smart meter raw and processed data (labeled as pseudo-measurements) sampled at low rate (5 to 20 times per hour) are combined with other measurements coming from regular SCADA systems an other smart grid sensors capturing the data each few seconds. The limitations of the conventional state estimation techniques are presented and the need of using different time scales for estimate the state of medium voltage distribution grids is evaluated. Again, the SM data are used in an aggregated way in a 100 buses, 15kV network.

It must be remarked that there is not any relevant study providing a comprehensive characterisation of the SMs measurement errors in domestic applications. In most of the cases, the analysis of the measurement error is carried out with aggregated data. In [11], the authors obtain the uncertainty bounds of the data provided by the SMs aggregated at substation level. They use real smart meter data sampled each 15 minutes for simulating loads in the IEEE low voltage European test feeder. In [14], a very sophisticated cloud based solution for performing state estimation in low voltage systems is proposed, but again, the data are aggregated at the concentrator level and the classical state estimation approach is used considering that the aggregated errors provided by the smart meters are normally distributed.

The work presented in [13] is one of the few studies in which a test to determine weather or not the data set obtained from the advanced metering infrastructure follows a normal distribution is carried out. In this case, the test is made with the data obtained from one of the buildings of the British Columbia University campus with a minimum consumption of 40kW and a maximum consumption of nearly 130kW. For this reason, the conclusions cannot be extrapolated to a domestic consumption profile. However, the paper introduces the concept of the out-of-date (OD) signals for characterising the error of the measurements provided by the SMs. A similar approach will be followed in this paper to analyse the errors of the domestic measurements and asses the viability of using the classical estate estimation theory to low voltage systems

Author email: arboleyapablo@uniovi.es

This work was partially supported by the Spanish Ministry of Economy and Competitivity under Grant MINECO-17-ENE2016-77919-R (CONCIALIATOR Energy conversion technologies in resilient hybrid AC/DC distribution networks). and by the Government of the Principality of Asturias under grant FC-15-GRUPIN14-127.

with non-aggregated smart meter data.

The structure of the paper will be as follows. In section II, the authors will state the very basics of the conventional state estimation theory and emphasising the need of normally distributed measurement errors. Section III will describe the data set used for this study. The concept of OD signals for characterising the error will be explained and applied to the data set. Section IV analyses the measurement errors by means of the Anderson-Darling test [13] and the conclusions about the probability density function (PDF) of the errors and the influence of the SMs set-up over these PDFs will be presented. In section V, in order to strengthen these conclusions, different power flows will be carried out in the European Low Voltage test feeder considering different SMs set-up. In the last section the general conclusions and recommendations for future works will be stated.

II. Classical state estimation approach

The classical state estimation theory is based on the maximisation of the joint probability density function of a set of measurements $f_m(\mathbf{z})$, where $\mathbf{z}^T = (z_1, z_2, ..., z_m)$ represent a vector containing the m measurements of the system [15]. Function $f_m(\mathbf{z})$ represent the probability of having a specific set of measurements and can be calculated as the product of the different probability density functions for each measurements if all measurements are independent.

$$f_m(\mathbf{z}) = f(z_1) \cdot f(z_2) \cdots f(z_m) \tag{1}$$

Assuming that the measurement errors are normally distributed with zero expected value and they are independent, maximising the joint probability density function ($f_m(\mathbf{z})$) is equivalent to minimising the function $J(\mathbf{x})$ described below.

$$J(\mathbf{x}) = \sum_{i=1}^{m} \frac{(z_i - h_i(\mathbf{x}))^2}{R_{ii}} = (\mathbf{z} - \mathbf{h}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{z} - \mathbf{h}(\mathbf{x})) \tag{2}$$

Where the term h_i represent the function relating the measurement i with the state vector $\mathbf{x}$. $\mathbf{h}(\mathbf{x})$ can be expressed as $\mathbf{h}(\mathbf{x}) = (h_1(\mathbf{x}), h_2(\mathbf{x}), ...h_m(\mathbf{x}))^T$. The matrix $\mathbf{R}$ (covariance matrix) is a diagonal matrix in which the term R_{ii} represent the squared standard deviation of the measurement i.

The expression presented in (2) is used for instance in [10], [13], [14] and in many other references. It must be emphasised that this expression requires normally distributed data and this condition is not always corroborated.

III. Data set and errors calculation

The data set used in this work was obtained from the project ADRES-CONCEPT [16]. This data set provides data of 30 different households in Upper-Austria during 7 winter days and 7 summer days, providing a total of 420 household profiles. The ADRES project provides active power, reactive power and voltage sampled each second in each of the three phases. In the present work, only the active power has been considered using a fixed power factor and loads are assumed to be single phase.

The active powers sampled at 1Hz will be processed to emulate different SM set-up considering sampling times (Δt) of 10s, 15s, 30s, 45s, 60s and 300s. The concept of out-of-date (OD) signal presented in [13] will be adopted to calculate the error assuming that the total error of the measurement i can be calculated as:

$$e_i^{total} = e_i^m + e_i^{OD} \tag{3}$$

Where e_i^m represent the SM measurement error and e_i^{OD} represent the error derived from the load variation with the time. There are two main factors when calculating this error, the specific profile of the analysed load and the time interval (Δt) used for setting up the SM. It must be considered that during the sampling interval k the measurement obtained in the interval $k-1$ will be used.

In Fig. 1 a) the active power in the household 5 of the above-described data set sampled at 1Hz is represented in black. The measurement error (e_i^m) will be neglected. It must be considered that the Smart Meters accuracy class is 0.5 [11], so the error added by the OD component (e_i^{OD}) will be much higher than (e_i^m). In red, it is represented the OD signal considering a Smart Meter sampling time of 5 min. It must be pointed out that the OD signal for the interval k is the mean value of the real signal in the interval $k-1$, so the e_i^{OD} can be calculated as:

$$e_i^{OD}(t) = P_i(t) - P_i^{OD}(t) \tag{4}$$

Where $P_i(t)$ is the active power consumed in the instant t and $P_i^{OD}(t)$ is the OD signal in the same instant. Two levels of zoom have been made in Fig. 1 a) in order to analyse $P_i(t)$ and $P_i^{OD}(t)$ from 16:00 to 20:00 hours and from 18:15 to 18:30. In this last case, and because the SM sampling time is set to 5 minutes we have three different intervals labeled as $k-1$, k and $k+1$. The instantaneous error from 18:15 to 18:30 is also represented. The average value of this error in the above described intervals is respectively 0.01kW, 2.95kW and 1.93kW as it can be observed. In two of the three analysed intervals, the mean value of the error is far from being null. Figs. 1 b) to g) represent the OD errors with different SM sampling times, from 10 seconds to 5 minutes in the time interval going from 18:00 to 19:00 hours. As it can be observed, the higher the sampling time, the higher the amount of instants in which the error is far from zero. This conclusion could be considered as trivial. However, up to the date, no statistical studies has been made to determine how far from zero are these errors and if they are or not normally distributed. In the next section an Anderson-Darling test will be performed to the 420 household profiles considering 6 different SM meters sampling times, analysing more than $217 \cdot 10^6$ errors.

IV. Analysis of the meassurement errors

Anderson-Darling test is commonly used to determine whether or not a specific data distribution (F) follows a normal distribution with an empirical cumulative distribution function

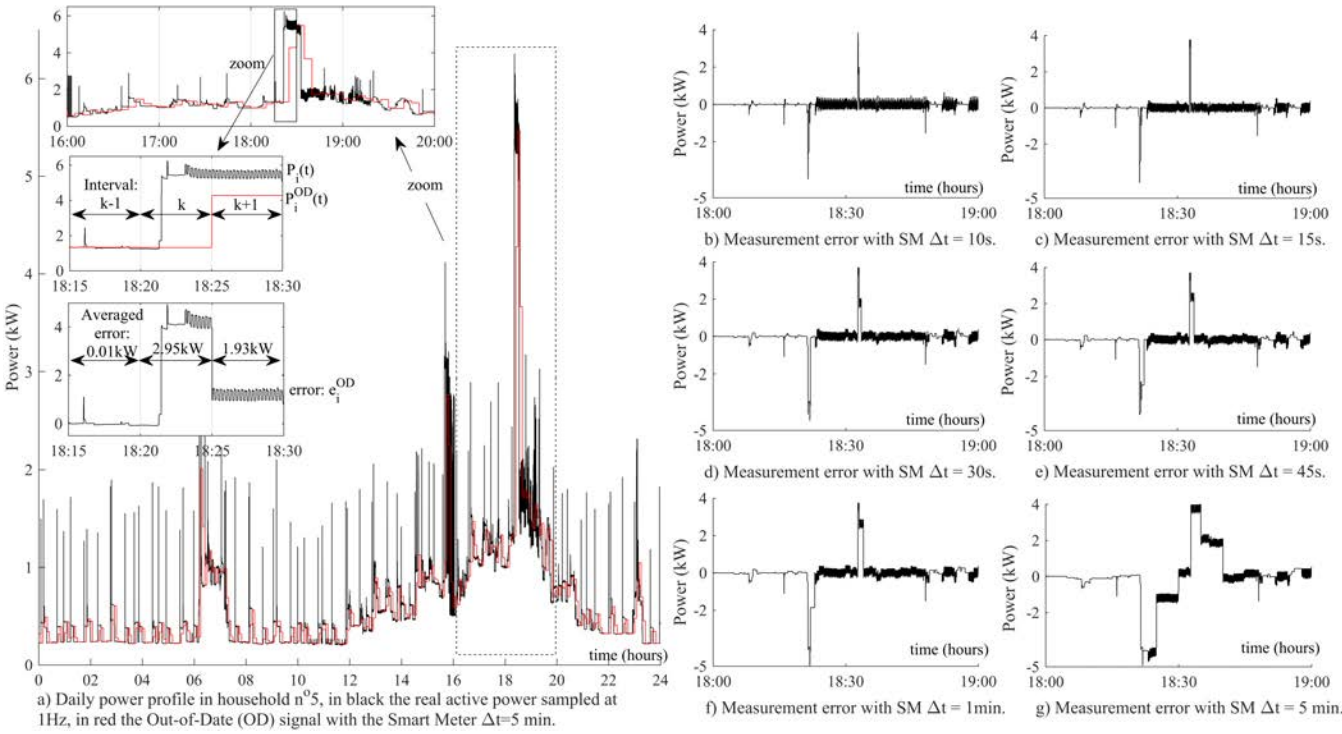

Fig. 1: Analysis of the active power in Household 5 and the measurement error using diferent Smart Meter set up.

(cdf) (F_n) calculating the distance between the two cdfs as follows.

$$p = \int_{-\infty}^{\infty} \frac{\left(F_n(x) - F(x)\right)^2}{F(x)\left(1 - F(x)\right)} dF(x) \tag{5}$$

This test was used for instance in [13] and it is going to be applied to the ADRES data set. We will consider 6 different sampling intervals for the SMs, 10s, 15s, 30s, 45s, 1min and 5min. During a specific sampling interval the measurement provided by the SM will be the OD signal ($P_i^{OD}(t)$) that will be constant. The Anderson-Darling test will be applied for each sampling interval individually, and will determine wether or not the instantaneous errors $e_i^{OD}(t)$ are normally distributed in that specific time interval. In addition, the average error per interval will be also presented. The so called "null-hypothesis" assumes that the data are normally distributed. As in [13], if the indicator labeled as p-value is lower than 0.05 the null-hypothesis can be accepted.

In Fig. 2, the summary of the tests made to the 420 profiles with different SM sampling frequencies is presented. In the first row of the figure, the p-value is represented for each SM set-up and for each interval in descending order. In the second row, the average value of the error in each interval is also representing but in ascending order per each SM set-up. Each column represent a different SM sampling time. In grey, the 420 profiles are represented and in red we can observe the average profile for each SM set-up. For instance, if we analyse the first column of the Fig. 2 (representing a smart meter sampling time of $\Delta t = 10s$), we can see that the average profile p-value is higher than 0.05 in the 73.5% of the intervals during the day. In those intervals it can be assumed that the error is normally distributed. In the second row (first column), we can observe that with the sampling period of 10s, the error in the average profile is lower than 100W in the 93.5% of the intervals during the day. The value of 100W was chosen as a limit for considering the average error neglectable.

As it can be observed in the Fig. 2, the higher the sampling period, the lower the percentage of intervals in which the error is normally distributed. For instance, analysing the 6th column (first row) of the figure ($\Delta t = 5min$), we can observe that less than 8% of the intervals have the error normally distributed for the average profile. In the case of the average error of the average profile, it is lower than 100W in the 69.10% of the intervals.

TABLE I: Percentage of intervals with normally distributed errors (Condition 1), percentage of intervals with average error lower than 100W (Condition 2) and percentage of intervals fulfilling both conditions at the same time.

Δt	10s	15s	30s	45s	1min	5min
Cond. 1(%)	63.9	53.2	38.1	29.8	24.3	4.3
Cond. 2(%)	96.9	96.7	95.8	95.4	94.6	86.8
Cond.1&2(%)	62.9	52.3	37.4	29.3	23.9	4.0

The Fig. 2 is very illustrative to analyse the behaviour of the p-value and the average error of the average profile separately. However it is interesting to know, how many of the intervals in which the error is normally distributed (Condition 1) have also an error lower than 100W (Condition 2). Table I contains the percentage of intervals fulfilling condition 1, condition 2 and both conditions and the same time, not for the average profile but for all profiles. As it can be observed, for a specific

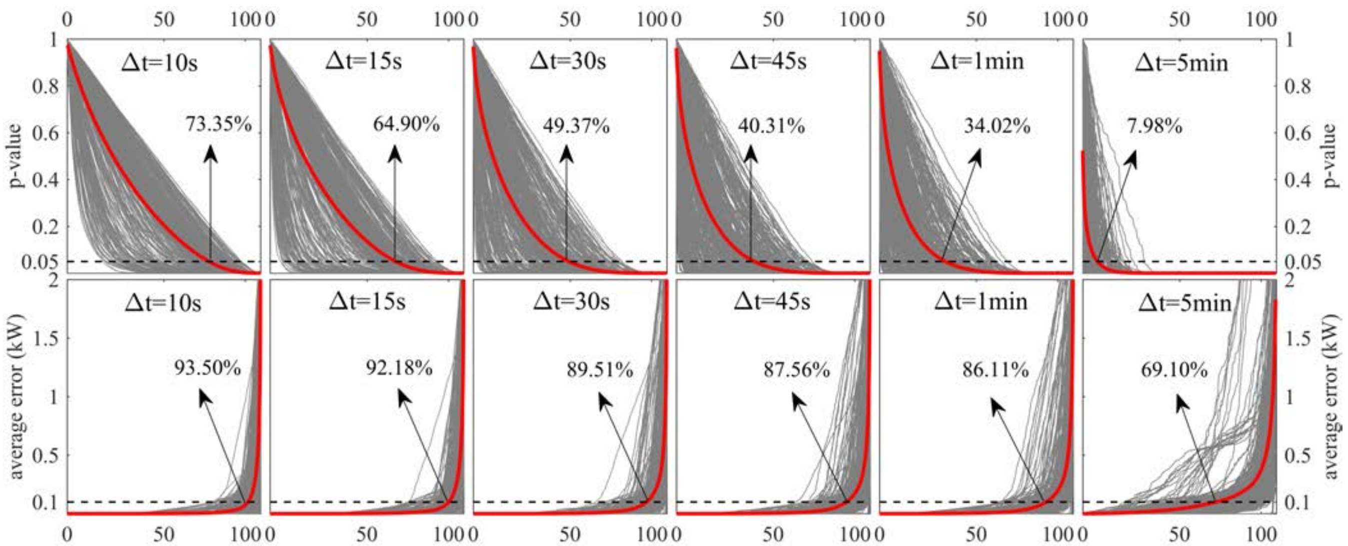

Fig. 2: Analysis of the measurement errors depending on the smart meters sampling time configuration. In the first row the p-value for Anderson-Darling test is represented, in the second row the average error in all the intervals is represented. In the horizontal axis the number of intervals is represented. For intance, with $\Delta t = 1min$, 100% will represent the 1440 intervals of 1 minutes

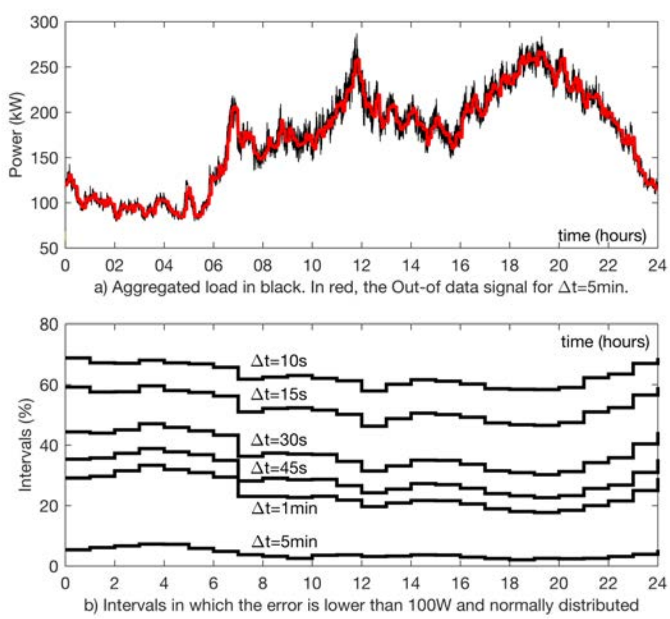

Fig. 3: Behaviour of the error in the different hours of the day depending on the day.

SM set-up the percentage of intervals fulfilling conditions 1 and conditions 1 and 2 is nearly the same. This means that for a specific interval, if the error is normally distributed, the probability of having a neglectable average error is nearly 100%.

However, during the day, the behaviour of the errors varies in a correlated way with the load of the system. In Fig. 3a) The aggregated power of the 420 profiles is represented (black), also the OD signal for the $\Delta t = 5min$ is represented in red. Fig. 3b) contains the percentage of the intervals that fulfil the two conditions described above at the same time (normally distributed errors and neglectable average error). As it can be observed, the curves for the different SMs set-up are nearly parallels. In addition, these curves are inversely correlated with the aggregated load of the system (Fig. 3a)). It could be concluded that for a given SM sampling time, the higher the load, the lower the probability of fulfilling both conditions 1 and 2 at the same time. For sampling periods longer than 30 seconds, the conditions for applying the conventional state estimation theory are fulfilled in less than 50% of the intervals no matter the time or the level of load of the system.

V. Load flow verification

In order to strengthen the conclusions extracted from the previous analysis, the IEEE European low voltage test feeder (Fig. 5) was loaded with 55 of the previous described profiles. The considered assumptions when solving the power flow problem are the ones described in [11]:

- All consumptions are monitored using Smart Meters.
- The smart meters measures are synchronised.
- The error of an individual measure can be neglected.
- The phase identification and the data management system work perfectly.

In Fig. 4a), the voltage at load 53 is represented. Fig. 4b) contains the voltage drop from the substation to the load 53. The results obtained when the system is loaded with the real data sampled at 1Hz are depicted in black. In blue and red, we can observe the results obtained using the OD signals of the smart meters for $\Delta t = 1min$ and $\Delta t = 5min$. It must be remarked that with $\Delta t = 5min$ set-up, the voltage profile is nearly flat during a normal day with voltage drops limited to 2V. The errors obtained when comparing the power flow results obtained with the signal sampled at 1Hz and the OD signals are depicted in Fig. 4 b), c), d), f), g), h). As it was expected, the higher the Δt, the higher the error.

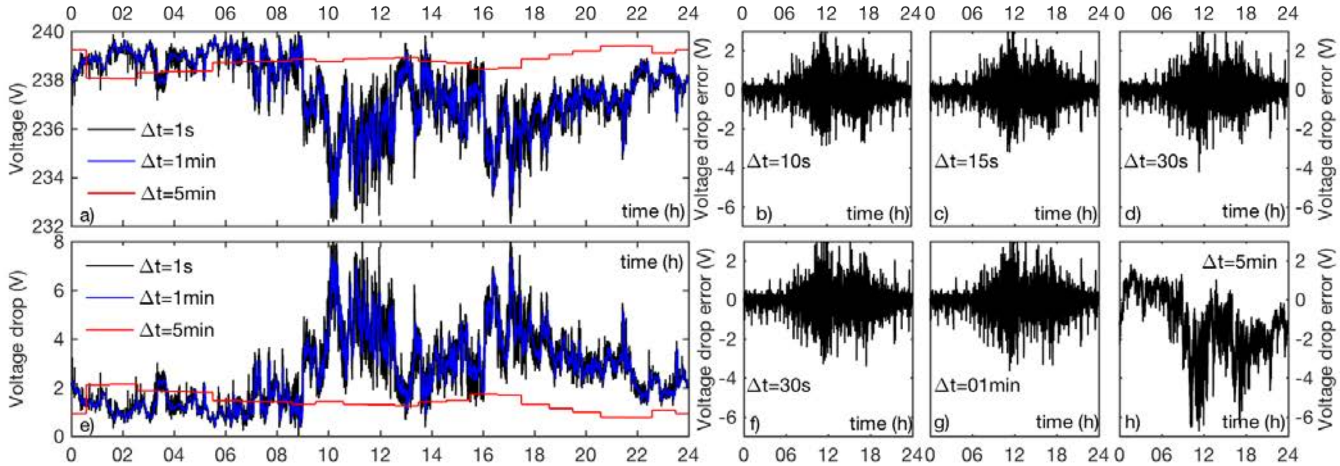

Fig. 4: Power flow results obtained using 1Hz sampled profiles as well as the OD signals with SM set-up of 10s, 15s, 30s, 45s, 1min and 5min. a) Voltage in the load 53 (See Fig. 5. e) Voltage drop from substation to load 53. b), c), d), f), g), h) Errors with the different OD signals with the above-mentioned SMs sampling times.

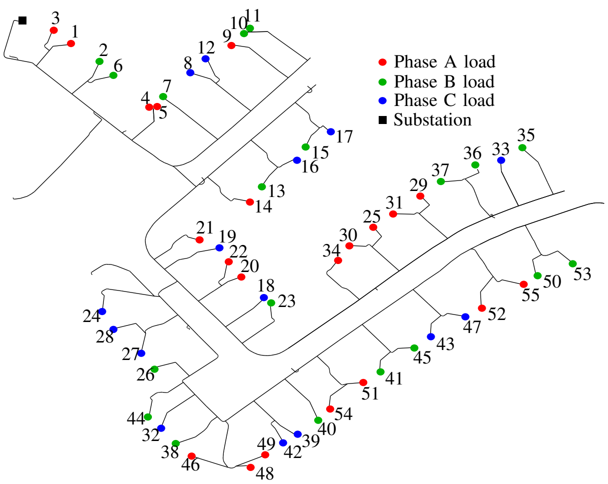

Fig. 5: IEEE European low voltage test feeder topology.

VI. Conclussions and Future Works

In this work, it was demonstrated that the non-aggregated data obtained from the smart meters (SMs) are not suitable for being used with the conventional state estimation techniques based on weighted least squares (WLS) due to the strong influence of the sampling time over the error and its probability density function (pdf). Even when the smart-meters are configured to send the data each 10s, the error in more than 37% intervals does not fulfil the requirements for being used with the WLS technique. Future works will consider the possibility of applying forecasting aided state estimation (FASE) or other techniques like Kernel density estimation especially designed for not-normally distributed errors.

References

[1] *Benchmarking smart metering deployment in the EU-27 with a focus on electricity*, European Comision, 2014.

[2] *Royal Decree 1110/2007, August 24^{th}. Unified regulation of measurement points in the electrical system*, Spanish Government, 2007.

[3] P. Kadurek, J. Blom, J. F. G. Cobben, and W. L. Kling, "Theft detection and smart metering practices and expectations in the netherlands," in *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Oct 2010, pp. 1–6.

[4] J. Peppanen, S. Grijalva, M. J. Reno, and R. J. Broderick, "Distribution system low-voltage circuit topology estimation using smart metering data," in *2016 IEEE/PES Transmission and Distribution Conference and Exposition (T D)*, May 2016, pp. 1–5.

[5] K. Soumalas, G. Messinis, and N. Hatziargyriou, "A data driven approach to distribution network topology identification," in *2017 IEEE Manchester PowerTech*, June 2017, pp. 1–6.

[6] D. Waeresch, "Topology fault detection in low voltage grids based on statistical analyses of smart meter data," in *2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*, Oct 2016, pp. 1–6.

[7] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, July 2015.

[8] H. Pezeshki and P. J. Wolfs, "Consumer phase identification in a three phase unbalanced lv distribution network," in *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Oct 2012, pp. 1–7.

[9] A. Primadianto and C. N. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, Sept 2017.

[10] V. D. Krsman and A. T. Sarić, "Verification and estimation of phase connectivity and power injections in distribution network," *Electric Power Systems Research*, vol. 143, no. Supplement C, pp. 281 – 291, 2017.

[11] F. Ni, P. H. Nguyen, J. F. G. Cobben, H. E. van den Brom, and D. Zhao, "Uncertainty analysis of aggregated smart meter data for state estimation," in *2016 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Sept 2016, pp. 1–6.

[12] A. Gómez-Expósito, C. Gómez-Quiles, and I. Džafić, "State estimation in two time scales for smart distribution systems," *IEEE Transactions on Smart Grid*, vol. 6, no. 1, pp. 421–430, Jan 2015.

[13] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, "Distribution system state estimation based on nonsynchronized smart meters," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov 2015.

[14] M. Pau, E. Patti, L. Barbierato, A. Estebsari, E. Pons, F. Ponci, and A. Monti, "Low voltage system state estimation based on smart metering infrastructure," in *2016 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, Sept 2016, pp. 1–6.

[15] A. Abur and A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*. Marcel Dekker, Ink, 2004.

[16] ADRES-CONCEPT, "(EZ-IF: Development of concepts for ADRES autonomous decentralized regenerative energy systems, project no. 815674). the project was funded by the austrian climate and energy fund and performed under the program ENERGIE DER ZUKUNFT." 2010.